

### Visualizing Public Opinion in Croatia Based on Available Social Network Content

Ševa, Jurica; Okreša Đurić, Bogdan; Schatten, Markus

Veröffentlichungsversion / Published Version  
Zeitschriftenartikel / journal article

#### Empfohlene Zitierung / Suggested Citation:

Ševa, J., Okreša Đurić, B., & Schatten, M. (2016). Visualizing Public Opinion in Croatia Based on Available Social Network Content. *European Quarterly of Political Attitudes and Mentalities*, 5(1), 22-35. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-459508>

#### Nutzungsbedingungen:

Dieser Text wird unter einer CC BY-NC-ND Lizenz (Namensnennung-Nicht-kommerziell-Keine Bearbeitung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:  
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de>

#### Terms of use:

This document is made available under a CC BY-NC-ND Licence (Attribution-Non Commercial-NoDerivatives). For more information see:  
<https://creativecommons.org/licenses/by-nc-nd/4.0>

## Visualizing Public Opinion in Croatia Based on Available Social Network Content<sup>1</sup>

Jurica Ševa

Advanced Computing Research Centre  
University of Sheffield, UK

Bogdan Okreša Đurić and Markus Schatten

Artificial Intelligence Laboratory  
Faculty of Organization and Informatics  
University of Zagreb, Croatia

*Date of submission: December 15<sup>th</sup>, 2015*

*Date of acceptance: January 23<sup>rd</sup>, 2016*

### Abstract

In the last decade advances of computer technologies have lead towards a technological reality where the line between information consumers and information producers is blurred. This technological omnipresence allows for unprecedented data creation capabilities. Based on various data sources, it seems humans have fully embraced data-generating activities. One such activity is using online social network applications, like Facebook or Twitter in almost all aspects of their lives. One of the main features of online social network applications is perceived freedom of speech, individuality and privacy, even though every application has some special features. Therefore, content generated using these services presents the public with interesting insights in private life of people and their attitudes towards public affairs. Social network applications are active the most during specific public events aimed at the massive public. Due to its brevity, ease of use and frequency, Twitter is an interesting social network application for research and analysis. Other than allowing almost exclusively short messages (up to 140 characters), a tweet (a Twitter post) can contain location of the message sender as well as a graphic to accompany the textual message. The textual part of the message may contain so called hashtags – keywords used for indexing and easy identification of a subject the message is related to. These hashtags allow us to group messages related to a specific event. Recent governmental elections held in Croatia were very popular amongst the Croatian Twitter community. Usage of hashtags allowed us to identify the right messages and thus most-used words to describe this event and potentially identify how people felt when talking, i.e. writing, about politics and the held elections. Furthermore, geolocation information, optionally embedded in a tweet, makes it possible to analyze which keywords were used in which parts of Croatia, all pertaining to the held governmental elections. Public opinion is only a few tweets away, but are the results similar to the election results? If tweet-based opinion can be constructed, does it differ from the real results? These and similar questions will be addressed in this paper.

**Keywords:** Twitter, social networks, web mining, opinion analysis, public data, Croatia, elections 2015



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/).

<sup>1</sup> Paper presented at the *European Conference on Political Attitudes and Mentalities*, 4th Edition, **ECPAM'2015**, University of Bucharest, December 2-3, 2015. MS PowerPoint Slide presentation of this paper available online on [ECPAM'2015](https://www.eupam2015.org/) website.

Corresponding Author: Bogdan Okreša Đurić, Doctoral candidate

Affiliation: Artificial Intelligence Laboratory, Faculty of Organization and Informatics, University of Zagreb

Address: Pavlinska 2, 42000 Varaždin, Croatia

e-mail: [dokresa@foi.hr](mailto:dokresa@foi.hr)

Copyright © 2016, Jurica Ševa, Bogdan Okreša Đurić and Markus Schatten

European Quarterly of Political Attitudes and Mentalities - EQPAM, Volume 5, No.1, January 2016, pp. 22-35.

ISSN 2285 – 4916

ISSN-L 2285 – 4916

## 1. Introduction

As demonstrated in some of our previous studies, social network services available online represent one of the best examples of globalization made real by the power of the Internet and provide us with new and unseen possibilities to study social phenomena (Schatten et al. 2015a; Schatten et al. 2015b; Ponjavić et al. 2010). Most certainly, leaders of this group of social network services, are those available publicly and are of most particular interest for social science studies. One such public social network system, which we chose for the analysis at hand, is the online microblogging service Twitter. This decision is based on several features of Twitter, including publicly available tweets (Twitter posts), an efficient and relatively easy data gathering process, as well as the quick nature and dynamics of Twitter (i.e. short posts of up to 140 characters, and streaming discussions about the presently relevant subjects). Undoubtful importance of Twitter in social activities performed on- or offline in the modern world is proved by the fact that on average 6.000 tweets are tweeted every second<sup>2</sup>. Furthermore, in only 7 years of its existence, the volume of tweets increased by six orders of magnitude, reaching a staggering number of 500.000.000 tweets per day in 2013 (see **Figure 1**). Although not intensely popular in Croatia, Twitter can spur some serious debate in geographic Croatia as Brautović demonstrates in his extensive analysis of anti-government protests in 2011 (Brautovic 2011; Brautovic 2012a, Brautovic 2012b). Nevertheless, analyses of Twitter in the Croatian blogosphere are very rare or non-existent.

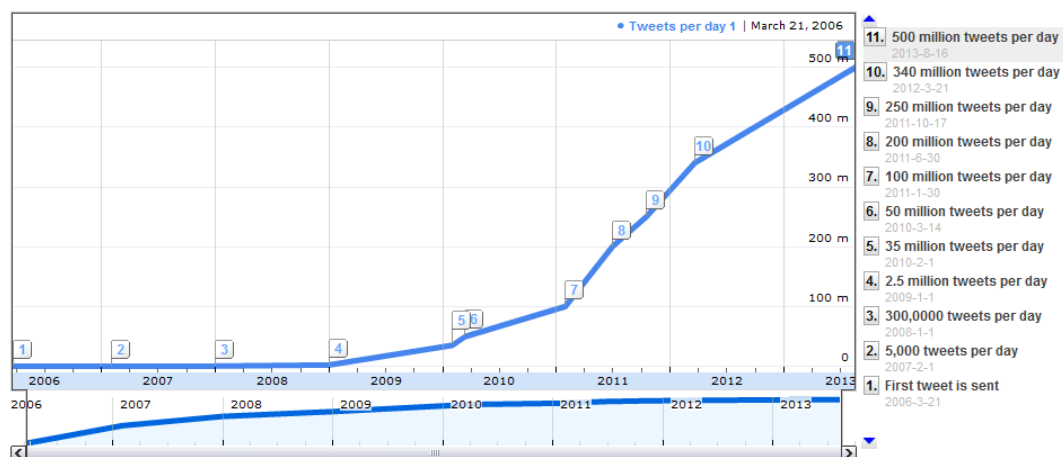


Figure 1.  
Historic Twitter volume, taken from [Internet Live Stats](http://www.internetlivestats.com/twitter-statistics/)

In the following we will try to fill this gap, and provide a current analysis of trends related to recent Croatian general elections. The main objective of this research is thus, to analyze the behaviour of Croatian people interested in politics that are skilled and interested enough to use the public space (especially Twitter) for expressing their opinion all in the context of the current political situation. We have gathered an extensive amount of social network data, that is analyzed herein.

<sup>2</sup> For more info see <http://www.internetlivestats.com/twitter-statistics/>

In the following we will try to answer some of the following questions: Does the tweet volume for each particular political party correspond with their respective voice count? What are the topics that are most used in the collected tweets? Can we gain insight into the concepts which Twitter users commonly connect with politics in the contemporary discourse? How interested are people to discuss politics or get engaged in politics-related subjects publicly?

The rest of this article is organized as follows: firstly in *Section 2* we provide a short outline of related research. In *Section 3* the data selection and collection process is described along with data filtering values, and data analysis and results are presented in *Section 4*. The results are interpreted and discussed in *Section 5*. *Section 6* concludes the research.

## 2. Related Research

Twitter and the content created by its users has been the subject of many published papers in the last decade. Meant for following trends, gaining access to large datasets, or predictive analyses, Twitter has often been viewed in the context of politics (Bekafigo 2013; Bermingham 2011; Brautovic 2011; Brautovic 2012a; Brautovic 2012b; Conover 2011; Dumitrica 2014; Groshek 2013; Tumasjan 2010; Park 2013; Sreekumar 2013; Tumasjan 2011; Younus 2011). Public opinion based solely on social networks has been in research focus as well, e.g. (Vesnic-Alujevic 2012; O'Connor 2010; Conroy 2012; Pak 2010). It has been argued that common tweets can present a researcher with valuable information when aggregated, albeit in the context of analyzing tweets about public health (Paul 2011).

Since its beginning in 2006, Twitter was most prominently introduced into the world of politics in 2008, and 2012, during the U.S. presidential elections (Groshek 2013). Results of the elections, and those of Twitter data analysis suggest that online social networks can be used to predict public opinion, yet are different than usual polls. Tumasjan et al. found that the mean absolute error of such a prediction is 1,65%. (Tumasjan 2011) The U.S. presidential elections of 2012 gave birth to as-of-then record breaking results of 327.452 tweets per minute.

Conducting a research with about 32.000 tweets on Irish General Election of February 2011, Bermingham and Smeaton conclude that, in general, Twitter may be used for predictions (Bermingham 2011). General Elections of May 2011 in Singapore were also followed up by an active Twitter public debate, and this online activity is argued as one of the major reasons for the opposition winning an unprecedented six seats in Parliament (Sreekumar 2013).

German federal elections of the national parliament held in September 2009 yielded 104.003 of tweets collected by Tumasjan et al. (2010). They concluded that political party ranking by tweet volume is almost identical to the rating by share of vote (Tumasjan 2011).

An argument of polls vs. social network data, analysis of both, and their results, is provided by O'Connor et al. (O'Connor 2010). Arguing in favor of Twitter, Bekafigo & McBride mention the ability to immediately observe user's behaviour in an activity (Bekafigo 2013). As one would expect, conventional polls require greater human effort, and can be costly and time-consuming, as opposed to online social networks data and the abundance of publicly available online information (Bermingham 2011).

A research based on Twitter data about 2010 mayoral race in Calgary, Canada, provides some interesting conclusions about the meaning of Twitter and other online social network platforms in general (Dumitrica 2014). Having risen a virtually unknown mayoral candidate to fame, social media is perceived as a communication medium between politicians and citizens that provides more equality and a more efficient information exchange than conventional media. The mentioned equality, and perceived directness, is due to seemingly no distinction of public and private in social media - an act of following a politician on a social

network service, sharing or commenting on political content (shared by friends or otherwise available), *“may coalesce into political engagement”* (Dumitrica 2014). Dumitrica further argues that a relatively small percentage of local population is actually involved in online social networks, yet there exists a belief that ‘everyone’ is present on social media.

*“To a certain extent, this discursive element encapsulates a genuine desire to be part of a community of friends, to live in a democratic society where politicians are accountable to citizens, and to have unmediated access to relevant information”*

(Dumitrica 2014).

Similarly, Park mentions several more motivators for using Twitter:

*“social participation through information exchange, forming a follower group based on interaction, information seeking and distribution, everyday conversation, checking public opinion, entertainment , and private expression”*

(Park 2013)

The democratic nature of Twitter is supported by the informal tone as well as ease of sharing, lowering barriers to access, and thus creating a perfect gateway for democratic media activism (Sreekumar 2013).

Interestingly, a number of papers argue that increased online presence and activity can increase traditional forms of political activity (Bekafigo 2013). Conroy, Feezell & Guerrero argue that increased online political activity is likely to bolster offline political participation, although without increased political knowledge amongst those involved (Conroy 2012).

In addition to public availability of everything posted on Twitter (Park 2013), thus increasing likelihood of interaction with strangers (Bekafigo 2013), equal possibility of posting is always present, since social status does not matter much on Twitter, and the service is widely available. Furthermore, it allows researchers to observe Twitter user’s behaviour while they are engaging in an activity, unlike polls. The mentioned public nature of Twitter, combined with fast pace of its usage,

*“opens the possibility for users to unknowingly be brought into the political process”*

(Bekafigo 2013)

On the other hand, large scale of Twitter and the mentioned fast pace, streaming nature, makes Twitter ideal for monitoring and analyzing events in real time (Conover 2011).

This last characteristic of Twitter is especially useful for analyzing the pace of change in events as they happen. In our case we have chosen to analyze the event of the Croatian general elections, not only because it is a politically important and interesting event, *per se*, but since the results of it were unpredictable and at the time of writing this paper are still not fully resolved. Due to a long lasting recession in Croatia, a large number of voters have decided to give their votes to a third, widely unknown option called cro. Most nezavisnih lista - Bridge of independent lists, as opposed to the two dominant coalitions in Croatia gathered around the two largest political parties the center-right HDZ - cro. Hrvatska Demokratska Zajednica - and the center-left SPD - Social Demokratska Partija each winning 59 places in the Sabor (the Croatian parliament) - not enough for each of them to form a majority government.

Most, as the 3rd runner-up with 19 places in the Sabor, found themselves in a situation where they had to decide who will run the state in the next four years. Having based their campaign on being the 3rd way, rejecting any possibility to form a coalition with any of the two larger parties (their leader Božo Petrov even publicly announced that he had signed a statement at the public notary's office that he will resign in case he forms a coalition with any of the two) they were under immense pressure from both parties as well as the public. In a twisted turn of events over a period of two months they engaged negotiations with both parties resulting in a 'contract of collaboration'<sup>3</sup> with HDZ, whereby a number of most representatives left the party in the process and formed independent clubs.

This social experiment continued with the insisting of Most to have an expert government, whereby at that time a publicly unknown financial manager from the pharmaceutical industry, Tihomir (Tim) Oreskovic was appointed as the mandatar, e.g. the candidate for the position of the prime minister, with no political background whatsoever.

A lot of other events in this elections also contributed to a wide public discussion, especially on social networks like Twitter. In the following we will try to gain a glimpse of that discussion by using a number of advanced web and data mining techniques.<sup>4</sup>

### 3. Data Acquisition

Every unit of Twitter data, called a tweet, consists of several elements. We have chosen to constrain the analysis towards the following fields: username (author of the tweet), date of creation, retweets (number of times the tweet was shared), favorites (number of users who favorited the tweet), and text (content of the tweet). A popular method of tweet categorization is by using hashtags, already mentioned above, which represent keywords used to group and identify key context of the tweet's content. Hashtags that we have chosen to follow are those identifying governmental elections in Croatia held in November 2015: #izboriHR (cro. izbori - elections, HR - Croatia's WWW domain appendix) which was chosen as the official hashtag of the elections by the Croatian national television (HRT - cro. Hrvatska Radiotelevizija), #izboriRH (similar as above, except RH is a Croatian acronym for cro. Republika Hrvatska - Republic of Croatia), #izbori2015 (meaning elections 2015), #politikaHR (cro. politika - politics).

Whilst using Twitter-data for analysis is of great use especially in the social sciences, there are a number of limitations that have to be mentioned here. Possibility of biased opinion of Twitter users in general is existent, since on the global level students are prevalent Twitter users<sup>5</sup>, akin to information technology, obviously willing to publicly disclose their thoughts. Croatia has a total of 3.788.788 voters<sup>6</sup>, and there are only 28.049 tweets generated by 2.838 users during data collection for the purpose of this study. Thus the results of the study cannot be directly applied for the whole voter set, but a subset of Twitter users that were interested in sharing their opinion on politics in the observed time period. Content found on Twitter is not verified to be the truth - every user is free to use their Twitter account according to their own preferences. Furthermore, users may be on a politically motivated mission of advertising a particular political party, thus affecting the available overall image. Selecting most appropriate hashtags for tweets filtering is a delicate job as well. The chosen four hashtags are definitely used to identify tweets on the subject of the Croatian governmental elections of 2015, yet the relatively small number of them limits

<sup>3</sup> The name 'contract of collaboration' was used to escape the unpleasant statement from the notary's office.

<sup>4</sup> For a detailed discussion on web mining techniques in social research please refer to (Schatten et al. 2015a and 2015b)

<sup>5</sup> For further information, visit *Most Common Keywords in Twitter Profiles* on <http://www.sysomos.com/inside-twitter>

<sup>6</sup> For further information, visit <http://www.izbori.hr/ws/index.html>



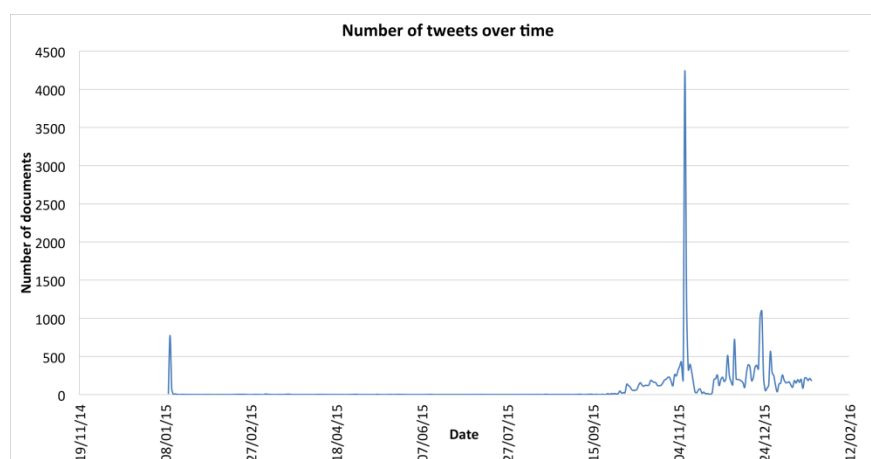
analysis possibilities. It might be better to collect tweets based on their geolocation, but since that is not a mandatory element in each message, our set of tweets with geolocation, amongst all the collected tweets, is minimal.

Yet, some of the results we have obtained are valuable and give an important insight into the semantics and dynamics of political topics in Croatia. Twitter is, as mentioned earlier, a highly valued medium for expressing opinion summed up in at most 140 characters, used rapidly and publicly, with data on recent and immediate events freely available for research and analysis. Croatian Twitter society, though still in development, might benefit from analyses such as this one, thus gaining popularity and building up. Furthermore, taking into account published research on the connections of Twitter and politics, conducting an experiment in a developing environment is of value for Twitter society.

After we identified hashtags which are mostly used to label our intended topic (i.e. #izboriHR, #izboriRH, #izbori2015, #politikaHR), tweets were saved in real time using a simple tool that was implemented for this study, which recognized tweets containing at least one of the chosen hashtags, and saved them in a repository. The use of Twitter's search function to find old tweets is limited temporally, i.e. one can search for tweets old at most about two weeks. This constraint introduced us to the problem of unavailability of historical data. Resolution to this problem was found in GetOldTweets,<sup>7</sup> a tool that allowed us to bypass some of these limitations. Data was collected from 21 Nov 2015 until 20 Jan 2016. During this period, as summarized in **Table 1**, a total of 28.049 tweets, made by 2.838 users, was collected, spanning January 2015 until 20 Jan 2015. It is obvious, from **Figure 2** which shows the amount of tweets through time, that the elections were a widely discussed topic on the day the elections were held, and that the topic remained popular afterwards, though not in such immensity.

**Table 1.** Statistical overview of the collected data

| Tweets | Users | Tweets/user | Tweets/day | Shared links |
|--------|-------|-------------|------------|--------------|
| 28049  | 2838  | 9,88        | 73,04      | 12264        |



**Figure 2.**

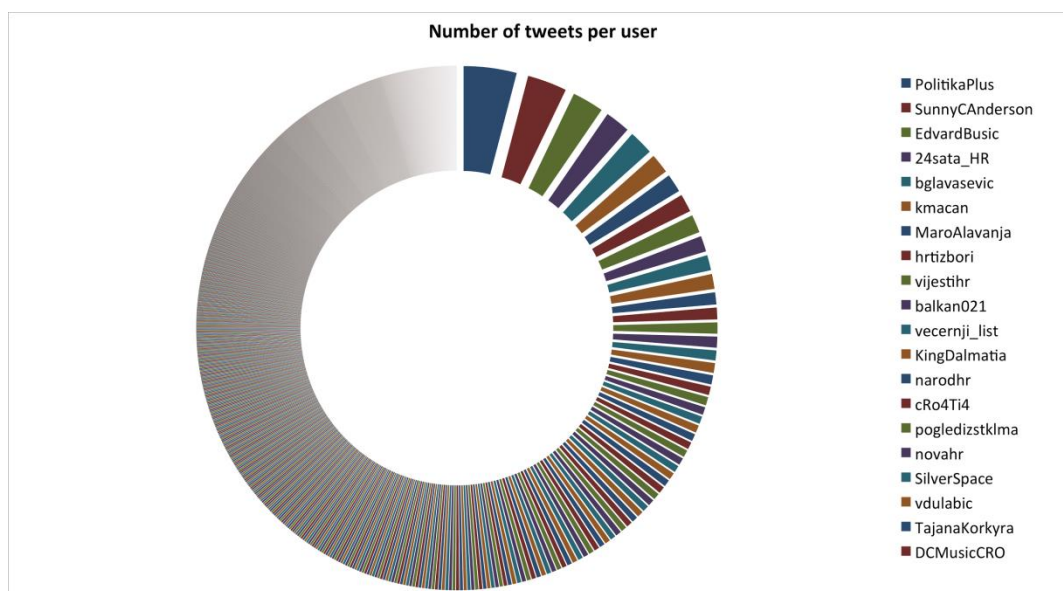
Topic popularity over observed time according to tweets volume per day

<sup>7</sup> Available at <https://github.com/Jefferson-Henrique/GetOldTweets>

#### 4. Analysis and Results

Initial data analysis of volume of tweets per day during the observed time period, as shown in **Figure 2**, clearly identifies a short period of time with increased tweets volume, and waning, although existent, popularity of the elections topic during the days following the elections held on 8 Nov 2015.

When observing **Figure 3**, according to the number of tweets per user during the observed period of time, several leaders, users with higher activity rate, can be identified. Since an average user created almost 10 tweets (as per **Table 1**), which corresponds to the global average Twitter user activity<sup>8</sup>, every user with more than 100 created tweets has increased activity by an order of magnitude, yet there are only five users (0,18%) with more than 500 created tweets. These users can be identified in **Figure 3**.



**Figure 3.**

Volume of tweets per user identifies several leading users

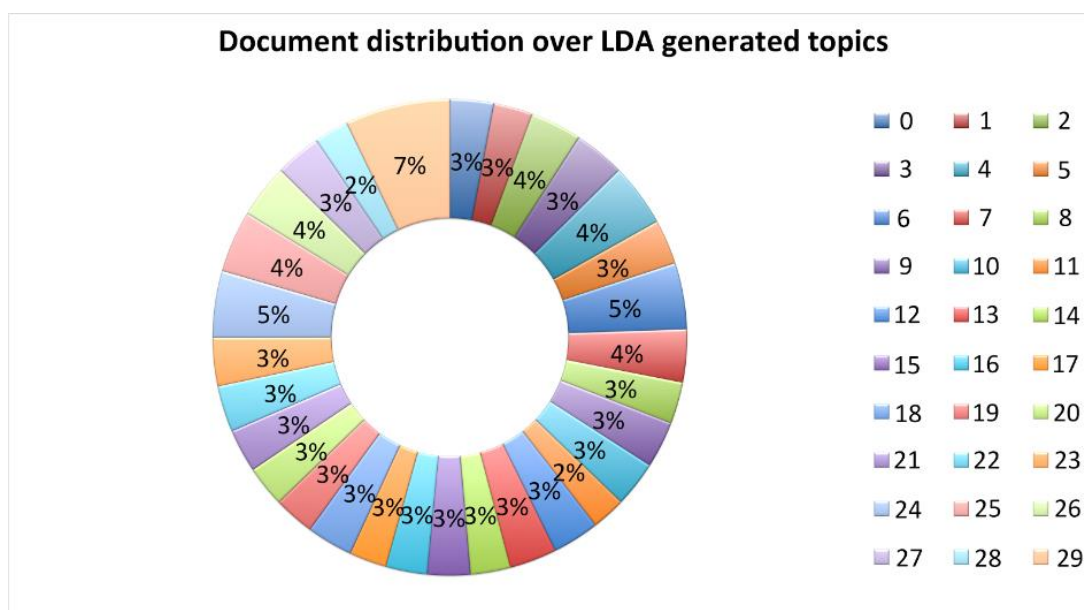
One of the objectives of this study was to create an overview of topics used in gathered tweets referring to Croatian governmental elections. In order to achieve this, topic modeling method was used, Latent Dirichlet Allocation - LDA (Blei et al., 2003) in particular, processed using a stemmer for Croatian<sup>9</sup>. This method is used to generate topics containing somehow connected words, based on their coexistence in the collected tweets. As depicted by **Figure 4**, 30 LDA topics have been generated (labeled 0-29). Words pertaining to a specific LDA topic, for the 7 LDA topics spanning the most documents (tweets), are shown in **Table 2**.

LDA topic 29 seems to be dealing with the obligations and intentions of the president which, according to procedure, has to provide the mandate to the option which has the majority in the Sabor. Words like cro. predsjednica (president), cro. nova (new), cro. vladu (government), cro. vodi (leads), and cro. sigurno (surely, secure) indicate this possibility.

<sup>8</sup> For further information, visit *Tweets/Day* on <http://www.sysomos.com/inside-twitter>

<sup>9</sup> For more information, see <http://nlp.ffzg.hr/resources/tools/stemmer-for-croatian/>





LDA topic 24 may deal with discussion about the future of what might happen after the elections and the installation of a new government. This is indicated by concepts like cro. će biti (will be, future tense), cro. neće (will not), cro. Hrvatsku (Croatia, accusative), cro. bit (will be), cro. sutra (tomorrow), and others.

**Table 2.** Most frequent words for 7 LDA topics with most documents

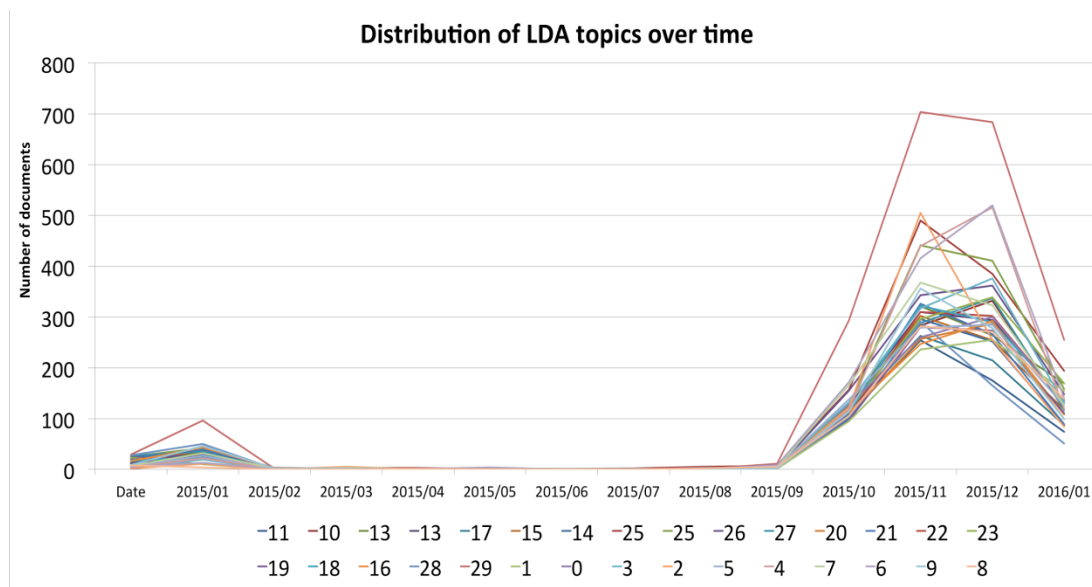
| LDA 29<br>("presidential obligations")  | LDA 24<br>("future")   | LDA 6<br>("political parties")  | LDA 4<br>("political leaders")  | LDA 25<br>("negative comments")   | LDA 26<br>("chances to stay in office")   | LDA 7<br>("TV comments")   |
|---|--|---|---|---|---|--|
| predsjednica, mjesta, nova, mora, vladu, ovih, nadam, tri, petrov, novu, lista, barem, puta, vodi, kgk, jedino, njih, sigurno | će, biti, koliko, neće, hrvatsku, sutra, bit, izborima, oko, građani, ako, već, svima, ovo, sve, puno, nama, osim, dobiti, kao | @nlmost, @hdz001, @sdphrvatske, oni, dan, pred, Ha, čak, novog, @d nevnikhr, pitanje, politički, žele, već, danas, moraju, @rtltelevizija, @no vahr | karamarko, milanović, nas, petrov, božo, kod, vise, zoran, dolazi, nije, sastanak, sve, valjda, ili?, iskreno, hoce, @youtube, kao, dr., samo | kad, sad, ste, sto, @novahr, jel, stranke, sve, ono, mene, manje, kako, znam, bili, kao, nema, govori, ima, koji, vidjeti | što, evo, nije, imaju, jos, nakon, dok, rekao, glasova, ka ko, izborne, mjesto, god, nikad, koju, zanima, kraju, mila novic, najveći, uopće | tko, @hrtvijesti, ima, nema, bez, ljudi, zna, netko, toliko, možda, izgleda, prvom, ako, [uživo], @kmacan, koji, pogledajte, kandidata, bar, ali |

LDA topic 6 may be identified as being about political parties and media (conclusion based on inclusion of numerous keywords identifying said organizations: NLmost, HDZ, SDPhrvatske, Dnevnik, NovaHR, etc.) and might deal with various questions (cro. pitanje - question) that were asked from the public.

Concerned with people, LDA topic 4 uses keywords representing names or surnames, e.g. Karamarko, Milanović, Petrov, Bozo, Zoran, etc. reflecting the leaders of the three most prominent political options in this elections.

LDA topic 25 might deal with commentaries on Tweets or articles from media (especially private television and portal NovaTV) with probably a negative sentiment. Most identified words are various adverbs that indicate comments on some other topic, e.g. cro. kad (when), cro. sad (now), cro. što (what), cro. jel' (is it), cro. sve (all), cro. ono (that), cro. manje (less), cro. kako (how), cro. bili (was), cro. kao (like), cro. nema (hasn't) etc. They might likely be short commentaries on some articles, dismissing their content, usual to political shills from various parties used to discredit opposite views.

LDA topic 26 seems to deal with discussing the chances of the current government to remain in power. Words including cro. evo (here, here are, behold), cro. nije (hasn't), cro. imaju (they have), cro. još (still), cro. nakon (after), cro. dok (until), cro. rekao (said), cro. glasova (votes), cro. izborne (electoral), cro. mjesto (place), cro. nikad (never), cro. kraju (in the end), as well cro. milanovic (SDP leader Zoran Milanović) indicate this.



**Figure 5.**

Popularity of LDA topics over time, on monthly basis (01/2015 - 01/2016)

LDA topic 7 may deal with commentaries on various TV related material, especially related to the Croatian national television (HRT). This is indicated by various concepts including @hrtvijesti (official HRT Twitter profile), cro. izgleda (looks like), prvom (first, first programme in slang), cro. uživo (live), and cro. pogledajte (look, watch).

To represent changes in LDA topics over time, dynamic topic modeling was used. This method allows us to determine changes in popularity, or numbers of documents (tweets) of a certain topic over the

observed time. The mentioned observation, on the other hand, indicates context in which the observed tweets were created, and how the context changed over time, determining volume of tweets of each topic at any given point in the observed time period. As depicted on **Figure 5**, only one LDA topic is the most popular one during the observed time period's visible user activity ("presidential obligations" from **Table 2**). Time of the elections, with the aftermath, is a time of turbulence for topic popularity though, with several topics changing volume of their respective tweets.

Though words are classified by topics, and their popularity and tweet volume is determined with the aforementioned methods, analysis of concepts mentioned in each tweet is done using the entity recognition method. Entities (keywords) are thus isolated and counted exclusively using hashtags as keyword identifiers.

The extracted keywords represent concepts which users thought would aid in defining context and content of their tweets. Keywords identified in this research, based on the collected data, are shown in form of a word cloud in **Figure 6**. Other than the word *cro. izbori* (elections), which is expected to be one of the most common, it is obvious the elections were governmental due to words *cro. sabor* (parliament) and *cro. vlada+rh* (government + RH - Republic of Croatia); held in Croatia - identified by words *cro. hrvatska* (eng. Croatia), *HRT* (Croatian national television), and *Croatia*; as well as that political parties are involved indicated by the words *HDZ*, and *cro. hrvatska raste* (Croatia is rising, name of the coalition formed around SDP), *cro. Živi Zid* (a marginal political party lead by relatively young people and based on activism).



**Figure 6.**

Generated word cloud of hashtags used after removing query hashtags

When 'tweeting' (creating tweets), user may incorporate in the content of their tweet a reference to another Twitter account using the sign @ immediately followed by the targeted user's username. Analyzing the occurrence of such mentions indicates targeted users for tweets that included user reference, thus allowing us to analyze the most mentioned users, hence their popularity. **Figure 7** shows most referenced

Twitter accounts in the observed collection of tweets. It is indicative that the three most mentioned accounts are political parties.

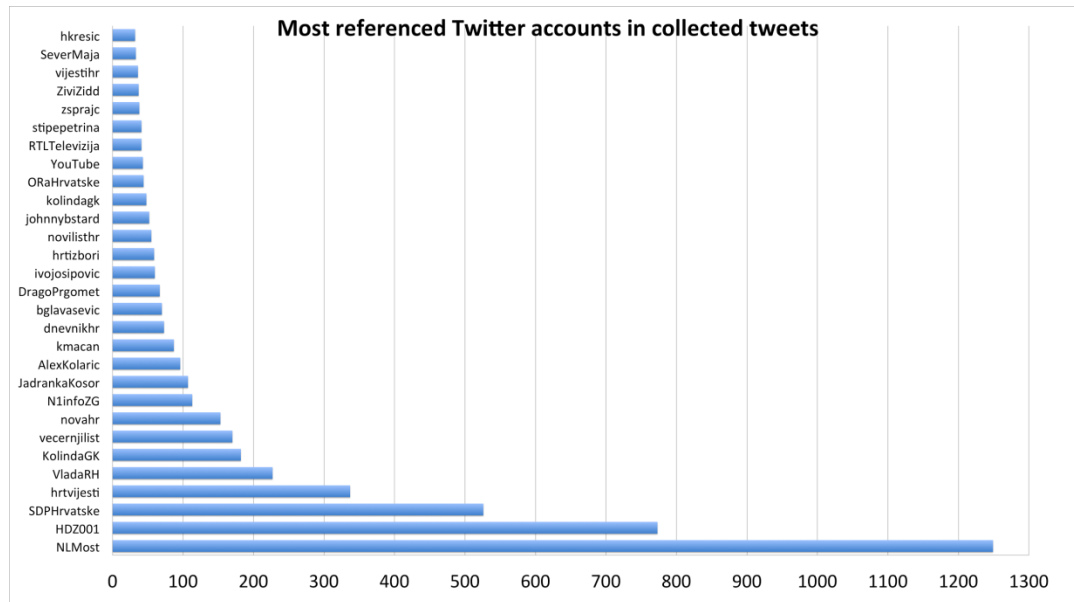


Figure 7.

Most referenced Twitter accounts in the observed collection of tweets

## 5. Discussion

Statistical analysis of the gathered data provided us with an insight into Croatian Twitter community's activity in the context of politics, especially with regards to the Croatian governmental elections of 2015. Results depicted in **Figure 2** indicate that the observed topic had uneven popularity during the observed period of time - naturally the elections were most popular at the day of the elections (8 Nov 2015), along with only a couple of days before and after the event. Time period immediately after the elections were held fell silent since people needed time to comprehend the results, and further events. Popularity of the subject rose again several days after the elections, which corresponds with the real-life event of forming political alliances for the new government - the situation was dire, since neither of the two most popular political options could form the government on their own.

Observations based on the volume of tweets per Twitter user, depicted in **Figure 3**, indicate that among the top five active users, there are only two representatives of the media. These results, coupled with the fact that amongst the top 20 active users, almost half are people, indicate that organizations, even though interested in advertising using the social media, and being involved in the Twitter world, are sometimes surpassed in their intent by private users.

By analyzing the 7 most active conversations, we have identified the main topics using LDA. Colloquially, we have labeled them (1) "presidential obligations", (2) "future", (3) "political parties", (4) "political leaders", (5) "negative comments", (6) "chances to stay in office", as well as (7) "TV comments".

Word cloud containing the most frequent hashtags (excluding those used to identify the tweets of interest for this research), shown in **Figure 6**, hides some interesting information. Summing some of the occurrence numbers, gives us an interesting insight into the context used by Twitter users: one political

option is described using keywords “hrvatskaraste”, “sdp”, and “milanovic” (name of the option, leading party of the option, and it’s leader’s name, respectively), while the other is described using keywords “domoljubnakoalicija”, “hdz”, and “karamarko” (same meanings as before). Summing their occurrence factors (numbers in **Figure 6**) gives us 19 points for the first option, and 20 points for the other option. These numbers correspond to the official election results (37,09% of votes for the first, as opposed to 39,07% for the other political option, whereas the rest is divided by political parties of smaller scale<sup>10</sup>).

The conclusion just mentioned is backed up by data represented in **Figure 7** where, when accounts of various political parties are filtered out, judging by the number of references of each of them, the situation is similar to that described above. The interesting detail to note is extremely high reference factor of the political option with username NLMost - this specific political coalition gained in popularity and significance after the elections were held, since they include enough politicians to turn the outcome of the elections in favor of one or the other larger-scale political options. Their immense popularity, as per **Figure 7**, is thus caused by their activities in the post-election period.

## 6. Conclusions

Analysis of Twitter data can give some important insights into public opinion about a certain event. Political events, such as elections mentioned several times through the course of this paper, may cause a surge of tweets, which are useful for constructing an image of public opinion about the included political options, predicting outcome of the event, or identifying key concepts used by the public, thus utilizing the discovered key concepts and using them to influencing the public. Predicting election results may not be a straightforward process though, since many elements have influence on popularity of a certain political option, and there are several different methods of measuring popularity of included political entities - two are covered by this research: hashtag occurrence and Twitter user referencing.

Even though Twitter is a rather praised source of data on various activities, as is evident from *Section 2* of this paper, which is mostly concerned with politically related data available on Twitter, some prerequisites for quality data collection have been identified through the course of this research, such as rather loose hashtag selection. The dataset used for this research is lacking in tweet volume, content, and number of hashtags leading to data skewness, and constrained methods of analysis. Yet, such is the state of Twitter society in Croatia.

Analyzing Twitter data, for political or any other purpose, certainly is not a costly and time consuming activity. Analysis is performed upon a rather structured set of data, data collection process is easy and automated once it had been set as intended, and the content published on Twitter is publicly available. Where a variety of content’s authors is needed, and desired data is already publicly available, Twitter is superior when compared to physical polls.

## Acknowledgement

This work has been supported in part by the Croatian Science Foundation under the project number 8537.

<sup>10</sup> For more information, explore <http://www.izbori.hr/ws/index.html>

## References

- Bekafigo, M. A., & McBride, A. (2013). Who Tweets About Politics?: Political Participation of Twitter Users During the 2011 Gubernatorial Elections. *Social Science Computer Review*, 31(5), 625–643. <http://doi.org/10.1177/0894439313490405>
- Bermingham, A., & Smeaton, A. F. (2011). On Using Twitter to Monitor Political Sentiment and Predict Election Results. In *Sentiment Analysis where AI meets Psychology (SAAIP) Workshop at the International Joint Conference for Natural Language Processing (IJCNLP)* (pp. 2–10). Chiang Mai, Thailand.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3, 993–1022.
- Brautović, M. (2011). Reporting of Croatian daily newspapers on 2011 antigovernment protests. *Kultura*, (132), 315–325.
- Brautović, M. (2012a). Utvrđivanje utjecaja internetskih aktivista: Upotreba Twittera u prosvjedima protiv hrvatske vlade. *MEDIANALI-znanstveni časopis za medije, novinarstvo, masovno komuniciranje, odnose s javnostima i kulturu društva*, 6(11), 1–12.
- Brautović, M. (2012b). Profiliranje Twitter-aktivista: Prosvjedi protiv hrvatske vlade. *Medijske studije*, 3(5), 51–63.
- Conover, M. D., Gonçalves, B., Ratkiewicz, J., Flammini, A., & Menczer, F. (2011). Predicting the Political Alignment of Twitter Users. In *2011 IEEE Third Int'l Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third Int'l Conference on Social Computing* (pp. 192–199). IEEE. <http://doi.org/10.1109/PASSAT/SocialCom.2011.34>
- Conroy, M., Feezell, J. T., & Guerrero, M. (2012). Facebook and political engagement: A study of online political group membership and offline political engagement. *Computers in Human Behavior*, 28(5), 1535–1546. <http://doi.org/10.1016/j.chb.2012.03.012>
- Cvetanova, G., Pachovski, V., and Bojadzievska, I. (2015) Republic of Macedonia and Citizen's Participation in the Digital Age: Where Do We Stand?, PowerPoint Presentation for ECPAM'2015, December 3, 2015, EPAM Conference Series.
- Dumitrica, D. (2016). Imagining engagement: Youth, social media, and electoral processes. *Convergence: The International Journal of Research into New Media Technologies*, 22(1), 35–53. <http://doi.org/10.1177/1354856514553899>
- Groshek, J., & Al-Rawi, A. (2013). Public Sentiment and Critical Framing in Social Media Content During the 2012 U.S. Presidential Campaign. *Social Science Computer Review*, 31(5), 563–576. <http://doi.org/10.1177/0894439313490401>
- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 122–129). Retrieved from <http://aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1536>
- Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and OpinionMining Alexander. In *Proceedings of the Seventh conference on International Language Resources and Evaluation* (pp. 1320–1326). <http://doi.org/10.1371/journal.pone.0026624>
- Park, C. S. (2013). Does Twitter motivate involvement in politics? Tweeting, opinion leadership, and political engagement. *Computers in Human Behavior*, 29(4), 1641–1648. <http://doi.org/10.1016/j.chb.2013.01.044>
- Paul, M. J., & Dredze, M. (2011). You Are What You Tweet: Analyzing Twitter for Public Health. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 265–272). <http://doi.org/10.1.1.224.9974>
- Pavlović, Z. and Todosijević, B. (2015) Public Understanding of Democracy in Serbia, Presentation PowerPoint for ECPAM'2015, December 3, 2015, EPAM Conference Series.
- Ponjavić, G., Schatten, M., Bača, M., & Jakelić, F. (2010). Personality Disorders in Croatia's Cyber Political Networks A Topic Map Study. In *Proceedings of the 21st Central European Conference on Information and Intelligent Systems*, Varazdin, Faculty of Organization and Informatics (pp. 33–39).
- Schatten, M., Ševa, J., & Okreša Đurić, B. (2015a). Big Data Analytics and the Social Web-A Tutorial for the Social Scientist. *European Quarterly of Political Attitudes and Mentalities (EQPAM)*, 4(3), 30–81.
- Schatten, M., Ševa, J., & Okreša Đurić, B. (2015b). An Introduction to Social Semantic Web Mining & Big Data Analytics for Political Attitudes and Mentalities Research. *European Quarterly of Political Attitudes and Mentalities EQPAM*, 4(1), 40–62.



- Sreekumar, T. T., & Vadrevu, S. (2013). Subpolitics and Democracy: The Role of New Media in the 2011 General Elections in Singapore. *Science Technology & Society*, 18(2), 231–249. <http://doi.org/10.1177/0971721813489458>
- Todosijević, B. and Pavlović, Z. (2015) Serbian Public Opinion on Democracy vs. Dictatorship in Serbia, Russia, Slovenia, Presentation PowerPoint for ECPAM'2015, December 3, 2015, EPAM Conference Series.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. In International AAI Conference on Web and Social Media. Retrieved from <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1441/1852>
- Tumasjan, a., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2011). Election Forecasts With Twitter: How 140 Characters Reflect the Political Landscape. *Social Science Computer Review*, 29(4), 402–418. <http://doi.org/10.1177/0894439310386557>
- Vesnic-Alujevic, L. (2012). Political participation and web 2.0 in Europe: A case study of Facebook. *Public Relations Review*, 38(3), 466–470. <http://doi.org/10.1016/j.pubrev.2012.01.010>
- Younus, A., Qureshi, M. A., Asar, F. F., Azam, M., Saeed, M., & Touheed, N. (2011). What Do the Average Twitterers Say: A Twitter Model for Public Opinion Analysis in the Face of Major Political Events. In 2011 International Conference on Advances in Social Networks Analysis and Mining (pp. 618–623). IEEE. <http://doi.org/10.1109/ASONAM.2011.85>

#### Open Access

This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.